ΛΜΙΛ
INFORMATICS PROFESSIONALS. LEADING THE WAY.

OXFORD

Research and Applications

Framework for identifying drug repurposing candidates from observational healthcare data

Michal Ozery-Flato [iD][1], Yaara Goldschmidt[2,†], Oded Shaham[2,‡], Sivan Ravid[1], and Chen Yanover [iD][2,§]

[1]Healthcare Informatics, IBM Research-Haifa, Mount Carmel Haifa, Israel and [2]Formerly Healthcare Informatics, IBM Research-Haifa, Mount Carmel Haifa, Israel

[†]Present address: K Health, Tel Aviv, Israel.

[‡]Present address: MeMed Dx, Haifa, Israel.

[§]Present address: KI Research Institute, Kfar Malal, Israel.

Corresponding Author: Michal Ozery-Flato, PhD, IBM Research-Haifa, Haifa University Campus, Mount Carmel Haifa, 3498825, Israel; ozery@il.ibm.com

ABSTRACT

Objective: Observational medical databases, such as electronic health records and insurance claims, track the healthcare trajectory of millions of individuals. These databases provide real-world longitudinal information on large cohorts of patients and their medication prescription history. We present an easy-to-customize framework that systematically analyzes such databases to identify new indications for on-market prescription drugs.

Materials and Methods: Our framework provides an interface for defining study design parameters and extracting patient cohorts, disease-related outcomes, and potential confounders in observational databases. It then applies causal inference methodology to emulate hundreds of randomized controlled trials (RCTs) for prescribed drugs, while adjusting for confounding and selection biases. After correcting for multiple testing, it outputs the estimated effects and their statistical significance in each database.

Results: We demonstrate the utility of the framework in a case study of Parkinson's disease (PD) and evaluate the effect of 259 drugs on various PD progression measures in two observational medical databases, covering more than 150 million patients. The results of these emulated trials reveal remarkable agreement between the two databases for the most promising candidates.

Discussion: Estimating drug effects from observational data is challenging due to data biases and noise. To tackle this challenge, we integrate causal inference methodology with domain knowledge and compare the estimated effects in two separate databases.

Conclusion: Our framework enables systematic search for drug repurposing candidates by emulating RCTs using observational data. The high level of agreement between separate databases strongly supports the identified effects.

Key words: drug repositioning, comparative effectiveness research, causal inference, Parkinson's disease, electronic health records

LAY SUMMARY

Drug repurposing is the quest to identify new uses for existing drugs. The gold standard for evaluating the effectiveness of a drug for a certain use (ie, treating a certain disease) is a randomized controlled trial (RCT), which is often costly and lengthy. Here, we present an easy-to-customize framework for identifying new uses for on-market prescription drugs by systematically analyzing healthcare databases, for example, electronic health records and medical insurance claims. Specifically, the framework computationally emulates, for each drug, a corresponding RCT: it provides an interface for extracting RCT components, including study population, patient attributes, and disease-related outcomes; then, it uses well-established causal inference methodology to estimate the drug effect on disease outcomes and its statistical significance (after correcting for multiple testing). We demonstrate the utility of the framework in a case study of Parkinson's disease (PD), where we evaluate the effect of 259 drugs on three PD progression measures. In this case study, we utilized two medical databases covering more than 150 million patients. The results of these emulated trials reveal a remarkable agreement between the two databases for the most promising candidates, strongly supporting the identified effects.

INTRODUCTION

Drug repurposing, or repositioning,[1] is the quest to identify new uses for existing drugs. It holds great promise for both patients and industry, as it significantly reduces the costs and time-to-market of new medications compared to *de novo* drug discovery.[2] To date, the most notable repurposed drugs have been discovered either through serendipity, based on specific pharmacological insights, or using experimental screening platforms.[2,3] To accelerate and increase the scale of such discoveries, numerous computational methods have been suggested to aid in drug repurposing (see reviews in refs[3–8]). For example, a popular approach, which can be applied to different data types, represents drugs and/or diseases as feature vectors (aka "signatures" or "profiles"), and measures the similarity between these entities or trains a prediction model for drug–disease associations.

In the healthcare domain, the term "real-world data" refers to information collected outside the clinical research settings; for example, in electronic health records (EHRs) or claims and billing data.[9] Such data offer important advantages in terms of volume and timeline span, alongside some inherent challenges such as data irregularity and incompleteness. Recently, real-world data have been increasingly leveraged for various healthcare applications.[10] In the context of drug repurposing, observational data are increasingly used to provide external validation to existing drug repurposing hypotheses. For example, Xu et al[11] used EHR data to validate the association of metformin with reduced cancer mortality. In contrast, there are far fewer examples for utilizing observational data to *generate* new drug repurposing hypotheses. Paik et al[12] derived drug and disease similarities from EHR data and then combined these similarities to score drug–disease pairs and suggest novel drug repurposing hypotheses. Kuang et al[13] leveraged patient-level longitudinal information available in EHRs and applied the Self-Controlled Case Series study design, widely used to identify adverse drug reactions,[14] to suggest new drugs that can control fasting blood glucose levels. Wu et al[15] employed a multivariable Cox regression model to assess the effect of 146 noncancer drugs on cancer survival in EHR data. Suchard et al[16] utilized EHR and claims databases to estimate the efficacy and safety profile of multiple first-line drug classes for hypertension treatment. They calculated the hazard ratios of the compared classes with Cox models, after stratifying or matching patients by propensity score.

Conducting a randomized controlled trial (RCT), the gold standard for validating the efficacy of a candidate drug, is costly and lengthy. To identify promising repurposing candidates, we propose a framework that *emulates* RCTs for on-market drugs using observational real-world data. We apply causal inference methodologies to correct for confounding bias in treatment assignment, treatment duration, and informative censoring.[17] Our framework is configurable, allowing for the specification of inclusion criteria, disease outcome, and potential confounders.

As a test case, we applied the described drug repurposing framework to Parkinson's disease (PD). To date, all drugs indicated for PD are approved for treating its symptoms and none was shown to slow the progression of the disease.[18] We emulated RCTs for hundreds of drugs, including PD indicated drugs, estimating their effect on three disease progression outcomes. To assess the robustness of our framework, we tested the agreement of the estimated effects across different causal inference methods and databases. We focus here on the methodological aspects of our framework and the means to validate its results. A discussion of the identified drug candidates for PD and their clinical validity appears in Laifenfeld et al.[19] To the best of our knowledge, our study is the first to demonstrate systematic emulation of RCTs in observational data for screening drug repurposing candidates.

MATERIALS AND METHODS

Each emulated RCT estimates the efficacy of a single drug by comparing patient outcomes in two cohorts: drug-treated patients (the "treatment cohort") versus controls; and correcting for biases between these cohorts, as well as biases related to treatment duration and incomplete follow-up. In the following sections, we describe the study design of the target trials, which largely follows the protocol in Hernán and Robins,[17] as well as our emulation framework. We focus on the parameters as customized to the PD study. These parameters, commonly defined in our framework with SQL queries, can be easily adjusted to other case studies.

Data sources

We analyzed two individual-level, de-identified medical databases. The IBM Explorys Therapeutic Dataset (freeze date: August 2017; "Explorys") included the medical data of over 60 million patients, pooled from multiple different healthcare systems, primarily clinical electronic heath records (EHRs). The IBM MarketScan Research Databases (freeze date: mid 2016; "MarketScan") contained healthcare claims information from employers, health plans, hospitals, Medicare Supplemental insurance plans, and Medicaid programs for ~120 million patients during the years 2011–2015. We note that

5.5 million patients (<10%) were known to be covered by both Explorys and MarketScan. As their content originated from different data providers, we consider them two separate resources and assume that the overlap in the derived patient cohorts and timelines is negligible.

Study design

Key dates

The beginning of the treatment, or its alternative in the trial, is termed the *index date*. We follow a new-user cohort design[17] and set the index date in our emulated RCTs to the first (ever) observed prescription date of the assigned drug. We refer to the observed time before the index date as the *baseline period* and use the information collected during that time period to determine whether a patient is eligible for the target trial. The time following the index date is termed the *follow-up period*, during which the effect of the drug is evaluated. In the PD case study, we set the follow-up period to two years. For each patient, we defined *end-of-treatment* as the end-date of the last prescription of the drug during the follow-up period. We considered patients as censored at the end-of-treatment. We set missing prescription duration in Explorys to three months, the modal value for prescription length in MarketScan data. We note that missing prescription durations can be more accurately imputed, for example, using each drug's statistics or with a regression model that additionally considers the patient's medical history. Finally, in the PD case study, we defined *PD initiation date* as the first PD diagnosis or the earliest levodopa (a drug typical to PD) prescription within the year preceding the first PD diagnosis. Since PD is likely present latently before the first diagnosis or prescription record, we retracted PD initiation date for all patients by additional 6 months. Technically, an earlier (presumable) PD initiation date increased the cohort size in our emulated RCTs, since we required the PD initiation date to precede the index date (see section Eligibility Criteria). Figure 1 illustrates the key dates in our study design.

Eligibility criteria

The target trials that we emulated focused on patients suffering from late-onset PD since early-onset patients present different clinical profiles.[20,21] We identified the late-onset patients in our data based on diagnostic codes (International Classification of Diseases, ICD, codes 9th and 10th revision) and required at least two PD diagnoses on distinct dates. Patients diagnosed with PD before the age of 55 were excluded from our study. To allow proper characterization of the patients in our trials, we further required an observed baseline period of 1 year. To ensure that all "recruited" patients have PD, we demanded that the PD initiation date precede the index date. The framework supports a relaxation of the new-user cohort design by setting the index-date to any prescription date having no prior prescription in some preceding time window (eg, a year).

Treatment assignment

For a given trial drug, the framework provides two possible settings for defining the control cohort. In the first setting, we use the Anatomical Therapeutic Chemical (ATC) classification system and set the alternative treatment to drugs from an ATC class of the trial drug, excluding the drug itself. Emulated RCTs comparing the target drug to all its encompassing hierarchical ATC controls may suggest mechanist explanations for the estimated effect or serve as sensitivity analysis.[19] In the second setting, the alternative treatment is a drug randomly selected for each patient from his/her list of prescribed drugs; additional criteria can be applied to limit this random drug set. See Discussion for the rationale behind these settings.

For both treatment and control cohorts, we demanded that the assigned treatment had at least two prescriptions ≥30 days apart. Finally, the framework excludes from the control cohort any patient with a prescription for the trial drug. The choice of control is configurable. In the PD case study demonstrated here, we tested both ATC-based and random-drug controls described above. In the ATC-based control, we used the second-level ATC class of each drug, noted as ATC-L2, which includes drugs of the same therapeutic indication. In the random-drug control, we considered drugs that are *not* indicated for PD.

Outcomes

The efficacy of a drug during the follow-up period is measured with respect to a patient-specific disease outcome, such as the occurrence
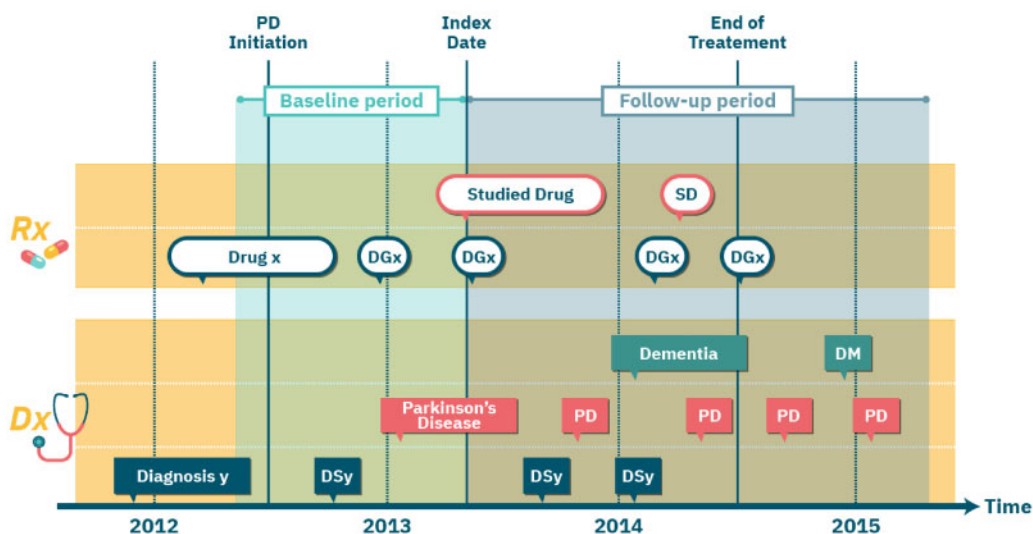

Figure 1. An illustration of the per-patient key dates in the study design of emulated randomized controlled trials. Each row corresponds to a certain type of medical event. Rectangles indicate diagnosis (Dx) events; ovals indicate prescription (Rx) events; event type is specified in the first (ie, leftmost) event in each row and then abbreviated (eg, "SD" in top row is the abbreviation for "Studied Drug").
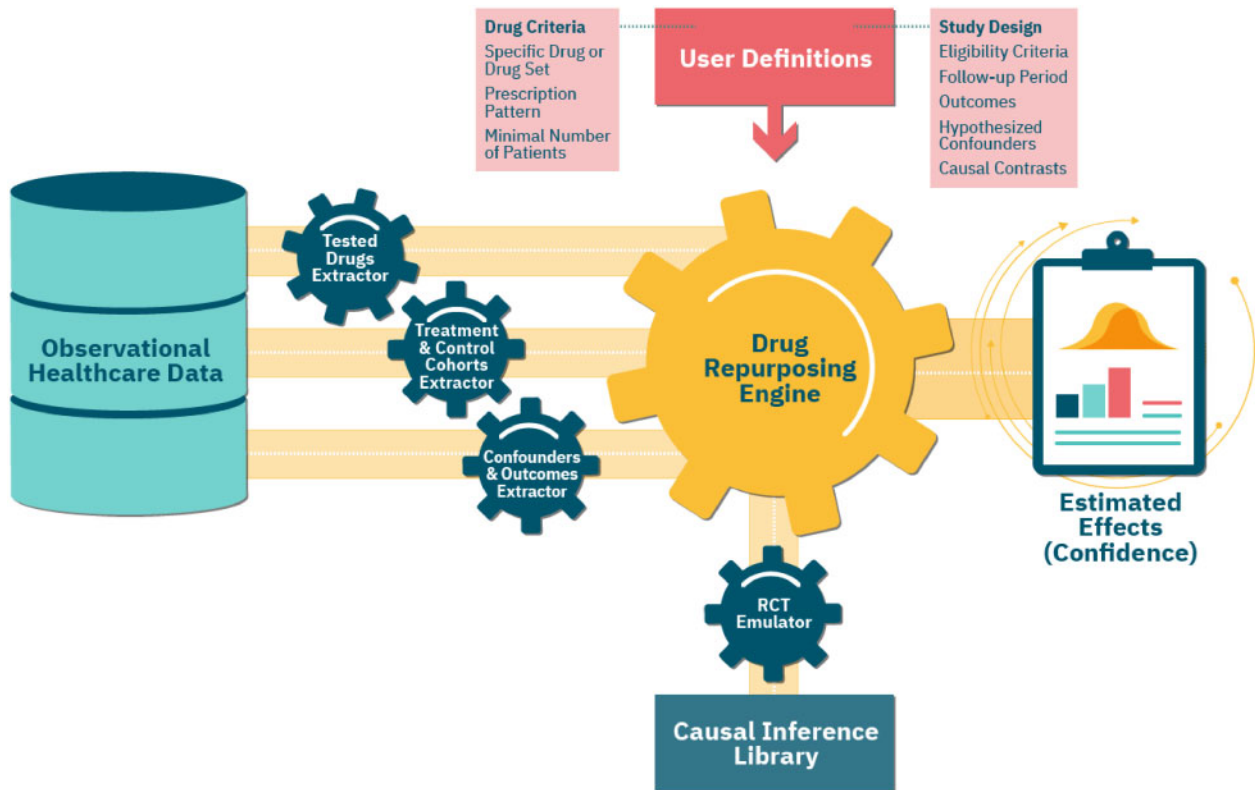
Figure 2. An overview of our framework's emulation pipeline and the underlying components. The central module is the Randomized Controlled Trial (*RCT*) Emulator, which orchestrates the entire process. First, the *Tested Drugs Extractor* identifies a list of repurposing candidates, based on the user-provided *Drug Criteria*. For each such candidate, using the input *Study Design* parameters, the *Treatment & Control Cohorts Extractor* assigns patients to the respective cohorts. The *Confounders & Outcomes Extractor* computes a baseline and follow-up attributes for patients in both cohorts. The *Drug Repurposing Engine* then instantiates an *RCT Emulator* for each candidate, which estimates its effect on disease outcomes in the treatment versus control cohorts, adjusting for the extracted confounders and using methods implemented in the *Causal Inference Library*.

of a disease-related event. In the PD case study, we defined a set of clinically-relevant events linked to the progression of PD along different axes:

- Fall: as a proxy to advanced motor impairment and dyskinesia.
- Psychosis: measuring progression along the behavioral axis.
- Dementia onset: measuring progression along the cognitive axis (and excluding patients with prior dementia diagnosis from the trial).

We used ICD codes to detect these events (see Laifenfeld et al.[19] for details).

The framework and its Causal Inference Library support effect estimation for continuous outcomes, such as lab test results. Note, however, that utilizing information from censored patients, for example, patients with incomplete follow-up period, is more challenging for such outcomes.

Hypothesized confounders
Confounders are variables affecting both the assigned treatment and the measured outcome, thus creating a "backdoor path"[22] that may conceal the true effect of the drug on the outcome. Causal effect estimation attempts to block these backdoor paths by correcting for confounders. Since, by definition, confounders influence the treatment assignment, they are computed over the baseline period. In the PD case study, our list of hypothesized confounders contained hundreds of covariates corresponding to: demographics (age and gen-

der), past diagnoses, prescribed drugs, healthcare providers' specialties, healthcare facilities utilization, and insurance types.

Framework for RCT emulation
Our configurable framework follows the study design protocol described above and automatically emulates a maximal number of RCTs using observational healthcare data. Figure 2 shows an overview of the framework and its RCT emulation pipeline.

Extracting tested drugs
We used the RxNorm standardized nomenclature to identify drug ingredients for each prescribed drug. Our framework tests all drug ingredients that satisfy the following conditions: (1) it is an active ingredient; (2) it is not part of over-the-counter medications, which may have limited coverage in our data; and (3) the number of patients in the corresponding treatment cohort is above a specified minimal value. In the PD case study, the minimum cohort size was 100 patients. The Tested Drugs Extractor module identifies all the drugs that meet these requirements (Figure 2).

Extracting treatment and control cohorts
For each emulated trial, our framework uses the feature extraction tool described in Ozery-Flato et al.[23] to extract the corresponding treatment and control cohorts, and to formulate and compute the values of the confounders and outcomes. In the random-drug control setting, the randomization process is shared by all trials, leading

to a large overlap between the control cohorts, and allowing a joint extraction of the confounders and outcomes in these cohorts.

Emulating an RCT

Below we provide a mathematical formulation of the estimated effects and elaborate on the steps our framework takes to evaluate them.

Let $P^{\text{trial drug}}(\text{outcome})$ denote the expected prevalence of patients experiencing an outcome event in an extreme scenario where *all* patients in the trial (ie, treatment and control cohorts) were assigned and fully adhered to the trial drug during a *complete* follow-up period. Similarly, let $P^{\text{alternative treatment}}(\text{outcome})$ denote the expected prevalence of the outcome for the analogous extreme scenario corresponding to the alternative treatment. Note that the evaluation of $P^{\text{trial drug}}(\text{outcome})$ and $P^{\text{alternative treatment}}(\text{outcome})$ may greatly deviate from $P(\text{outcome}|\text{treatment cohort})$ and $P(\text{outcome}|\text{control cohort})$, namely, the observed (uncorrected) outcome prevalence in the treatment and control cohorts, due to biases in treatment assignment, treatment duration, and loss-to-follow-up. The effect of the trial drug on the outcome is then measured by the difference.

$$P^{\text{trial drug}}(\text{outcome}) - P^{\text{alternative treatment}}(\text{outcome})$$

Alternative ways to measure the effect are the ratio and odds ratio of $P^{\text{trial drug}}(\text{outcome})$ and $P^{\text{alternative treatment}}(\text{outcome})$. The choice of effect measure is configurable and depends on the goal of the inference (but odds ratio is less preferred due to its non-collapsibility[22]). Procedure 1 estimates the effect of the trial drug on an outcome, as well as its statistical significance. The steps in this procedure are implemented in the Causal Inference Library module. Details are provided in the following section.

Procedure 1: RCT Emulation

Input: Patient data: assigned treatment, outcome, censoring time, observations of hypothesized confounders
Output: Estimated effect (and P-value) of treatment on the outcome

1: Identify major confounders from the list of hypothesized confounders by their association with the outcome
2: **repeat**
3:   Compute balancing weights for the treatment and control cohorts
4:   Set the state *is_balanced* to **true** if no major imbalance exists between the reweighted treatment and control cohorts; Otherwise set it to **false**
5:   **if not** *is_balanced* **then**
6:     Increase positivity likelihood by patient exclusion
7: **until** *is_balanced* **or** number of patients in the treatment or control cohorts is too small
8: **if not** *is_balanced*
9:   **return** "Cannot evaluate treatment effect"
10: **else**
11:   Estimate the treatment effect (using causal inference method)
12:   Estimate the P-value of the computed effect (using bootstrapping)
13:   **return** the estimated effect and P-value

Causal inference library

This module contains various methods to estimate the expected outcomes and causal effects from observational data. For event-based outcomes, as in the PD case study, it offers causal survival analysis methods that adjust for both confounding and selection bias due to incomplete treatment period. Below we provide a summary of the methods used in the PD case study.

Identifying major confounders. This step identifies major confounders within the set of extracted potential confounders by testing their association with the outcome[24] (Step 1; see also Discussion). First, it excludes features that are nearly constant (mode frequency > 0.99) in both the treatment and control cohorts. It then dichotomizes non-binary feature values into high/low using their median, and measures the association between the feature and the outcome using the following difference

$$P(\text{outcome}| \text{ high feature values}) - P(\text{outcome} \mid \text{low feature values})$$

For event-based outcomes, we computed this difference with Kaplan–Meier estimators and used bootstrapping to assess its statistical significance.[25] In the PD case study, we used a P-value ≤ 0.005 in all emulated trials to identify major confounders.

Generating balancing weights. To generate balancing weights (Step 3), we applied the popular method of inverse probability weighting (IPW) with stabilization,[26] and modeled treatment probability (propensity score) with logistic regression. To avoid large variance in the resulting estimands, we used weight trimming for percentile range 1–99%.[27,28]

Testing for imbalance. We tested the imbalance between two, possibly weighted, cohorts (Step 4) by computing the absolute standardized difference[26] for each identified major confounder:

$$d = \frac{|\bar{x}_{\text{treatment}} - \bar{x}_{\text{control}}|}{\sqrt{(s^2_{\text{treatment}} + s^2_{\text{control}})/2}}$$

where $\bar{x}_{\text{treatment}}$, $\bar{x}_{\text{control}}$ are the feature means in the two treatment groups, and $s^2_{\text{treatment}}$, s^2_{control} are the corresponding sample variances. We referred to the cohorts as *balanced* if for all major confounders $d \leq 0.2$.[29]

Increasing positivity likelihood. Emulating RCTs requires satisfying the positivity condition: each patient in the trial has a positive probability of receiving either the trial drug or the alternative treatment. A failure to find balancing weights (Step 5) may indicate a violation of this condition. To increase the likelihood that the positivity condition is satisfied (Step 6), we excluded patients whose propensity scores lay outside the overlap of the treatment and control cohorts.[30]

Effect estimation. The Causal Inference Library provides two different methods for estimating treatment effects (Step 11), based on: (1) balancing weights, and (2) outcome prediction. The former method estimates the expected outcome for each treatment, $a \in \{\text{trial drug, alternative treatment}\}$, using data reweighting. For event-based outcomes, we use a Kaplan–Meier estimator that reweights patients at each time unit to adjust for both confounding bias and informative censoring (the latter may be leading to a selection bias):

Table 1. Summary of trials in Explorys and MarketScan for the PD case study. Emulated trials correspond to drugs with balanced treatment and control cohorts in both Explorys and MarketScan (percentage out of the tested drugs is shown in parentheses). An identified drug's estimated effect reduces the prevalence of the corresponding PD outcome at FDR <0.05 in both Explorys and MarketScan.

Outcome	Control cohort	Explorys and MarketScan trials and results		
		Tested drugs	Emulated trials (%)	Identified drugs
Dementia	random-drug	223	183 (82%)	0
Dementia	ATC-L2	218	205 (94%)	2
Fall	random-drug	259	219 (85%)	0
Fall	ATC-L2	247	228 (92%)	1
Psychosis	random-drug	259	218 (84%)	0
Psychosis	ATC-L2	247	214 (87%)	1
Any outcome	random-drug	259	259 (100%)	0
Any outcome	ATC-L2	247	247 (100%)	4

$$P^a(\text{outcome}) = \prod_{t=1}^{\text{follow-up length}} \left(1 - \frac{\sum_{\{i:\ A_i=a\}} w_{t,\ x_i} 1[T_i^{\text{outcome}} = t]}{\sum_{\{i:\ A_i=a\}} w_{t,\ x_i} 1[T_i \geq t]} \right)$$

where i denotes a patient, A_i denotes the assigned treatment, T_i^{outcome} denotes the time of first outcome event, T_i denotes the minimum of censoring (ie, end-of-treatment) and outcome event times, and $w_{t,\ x_i}$ is the computed balancing weight in time t. In the PD case study, the time unit was one month (30 days).

The second method predicts the expected outcome for a treatment for each patient in the trial and then estimates the overall expected outcome for the treatment by taking the average:

$$P^a(\text{outcome}) = \frac{1}{n} \sum_i P[\text{outcome} X = x_i,\ A_i = a]$$

We provide complete details of these methods in the Supplementary Material. In the PD case study, we applied both methods for computing effects and analyzed their agreement.

P-values estimation. We evaluated the *P*-value of an effect (Step 12) using a bootstrap estimate of its standard error and assuming the distribution of a sample effect is close to normal.[25] To account for multiple testing, we controlled for the false discovery rate (FDR) using the method of Benjamini and Hochberg.[31] Adjusted *P*-values ≤ 0.05 were considered statistically significant.

RESULTS

We identified ∼106 000 patients in MarketScan and ∼89 000 patients in Explorys as eligible for our emulated PD trials. To get a notion of the differences in the patient population participating in the emulated trials in each database, we compared the random-drug control cohorts in these databases. This comparison revealed many similarities, such as the average age (∼75.5 years), percentage of women (43–45%), and the fraction of patients with public insurance (82–85%). In both databases, dementia was the most prevalent outcome during the 2-year follow-up (37–45%), followed by fall and psychosis (17–26% and 10–15%, respectively). There are also notable dissimilarities between the two databases; the most prominent is the average total patient time in database, which was more than twice as long in Explorys compared to MarketScan.

Table 1 provides statistics on the trials and results in the Explorys and MarketScan databases. Overall, we tested the effect of 259 drugs on psychosis, dementia, and fall in 1453 emulated trials, using different controls and databases. There are fewer trials with

an ATC-L2 setting due to smaller control cohorts or missing ATC. For most (82–94%) of the drugs, the RCT emulator successfully generated balancing weights in both databases (since confounders are selected based on their association with the outcome, balancing rates vary between outcomes). Despite the greater statistical power of random-drug control cohorts, we obtained more significant results using ATC-L2 control cohorts. Only 4 (0.3%) of the 1,453 trials ended with significant beneficial effects at FDR 5% by the two causal estimation methods and in both databases. These 4 trials involved 4 distinct drugs: rasagiline, zolpidem, azithromycin, and valsartan.[19] Rasagiline, which was shown to be significantly associated with a lower rate of dementia onset, is currently narrowly indicated for treating PD motor symptoms. Deeper analysis of rasagiline and zolpidem's effects and discussion of their potential mechanisms of action appear in ref.[19].

We next applied a meta-analysis of estimated effects and assessed the level of agreement between the two different causal inference methods, namely balancing weights and outcome models. We observed (Figure 3) strong and significant correlations between the effects estimated by the two causal inference methods (focusing on drugs where at least one of the estimated effects is significant at FDR of 5%). These correlations appear to be stronger in Explorys than in MarketScan. Importantly, significant effects by the two methods always agreed on effect sign (ie, beneficial vs. harmful).

Finally, to test the agreement between Explorys and MarketScan, we restricted the comparison to drugs that were shown to have a significant effect by the two causal estimation methods in both databases (Figure 4). The agreement between the estimated effects is remarkable, with a perfect match for effect sign, and near equivalence in the magnitude of the effects. A comparison of the corresponding uncorrected effects shows similar, though somewhat weaker, agreement between the two databases (see Supplementary Figure S1).

DISCUSSION

A correct inference of causal effects must involve a subject-matter expert, who is also aware of the data-generation process.[32,33] Our framework allows easy injection of domain knowledge into the implementation of the emulated trials, including: the formulation of cohorts, outcomes and hypothesized confounders.[23] There are caveats for the use of healthcare data in comparative effectiveness research, including inaccurate and incomplete information, which require in-depth understanding and careful interpretation of the data.[34] Specifically, there are potentially many sources of bias in
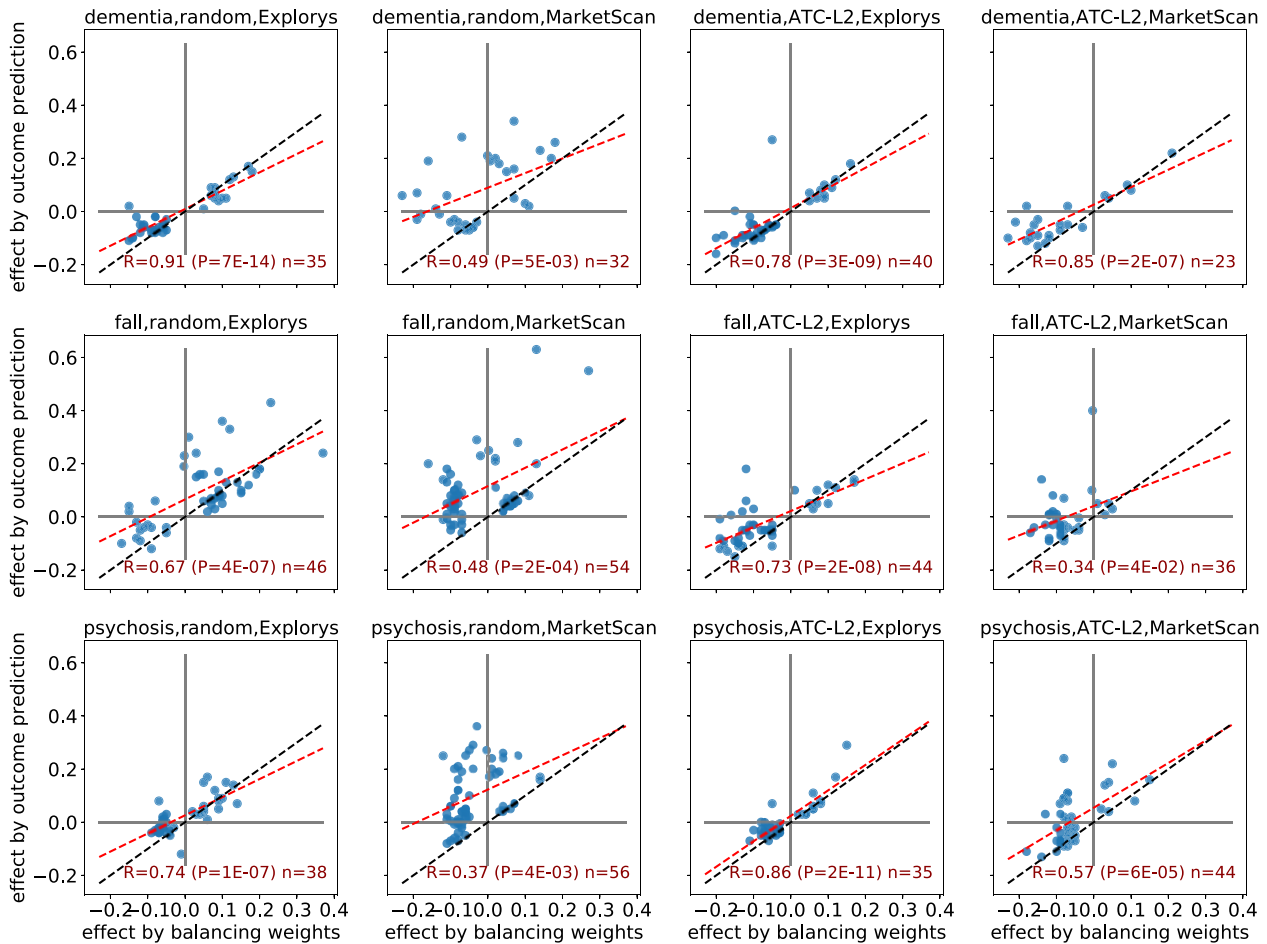
Figure 3. A comparison of estimated effects: balancing weights vs. outcome prediction. Each chart shows a different setting of the trial with respect to the outcome (dementia, fall, and psychosis; rows), control cohort (random and ATC-L2; left and right columns), and database (Explorys and MarketScan; alternate columns). Each point corresponds to a drug whose estimated effect was significant at FDR 5% by at least one of the two compared methods. The red line is the fitted least squares regression line; blue line indicates $y = x$.

healthcare data,[35] leading to a large set of hypothesized confounders, some not directly recorded in the data but can often be approximated using observed variables. Adjusting for a large number of confounders may result in non-positivity, high-variance estimates of the effect, and over-adjustment bias.[24,36] Our framework takes a combined approach by extracting a very large number of features suspected as confounders and then applying a confounder selection step (Procedure 1, line 1). The incorporation of a feature engineering tool,[23] makes our framework unique in the ease and flexibility of defining numerous potential confounders. We applied a strategy that identifies major confounders based on their statistical association with the outcome.[24] In the PD case study, we focused on monotone associations for non-binary variables, testing high vs. low variable values, but advanced examinations may consider non-monotone associations as well. Furthermore, other methods for confounder selection, for example, Vansteelandt et al. and Ertefaie et al.[37,38] may be considered as well. To the best of our knowledge, our drug repurposing framework is the first to correct for selection bias caused by associations of outcome-related variables with follow-up or treatment duration.

Another novel aspect of our framework is the optional use of a random-drug control, where patients in the treatment cohort are compared to other patients with the studied disease who start any different treatment. The variety of drugs in the random-drug cohort represents the "background" medications prescribed to patients with the disease. Therefore, the estimated expected outcome in this cohort approximates the "average" outcome in the entire study population. Random-drug control cohorts are relatively large, thus potentially increasing the statistical power of the emulated trials. Additionally, under this control setting it may be easier to compare the estimated effects of different drugs, as all drugs are tested against a similar set of alternative drugs. A potential caveat of the random-drug control setting is that the populations of the treatment and control cohort may be very different, thus increasing the risk for uncorrected confounding biases. Alternatively, the ATC-based control option allows the user to restrict the control cohort to patients initiating a more homogeneous treatment that is related to the trial drug. The higher the ATC level, the more closely related the pharmacological and chemical properties of the trial drugs and their alternatives in the control, potentially ensuring a greater resemblance, or match, between the patients in the treatment and control cohorts.[39] However, as the estimated effect is a comparative measure, it may be obscured when the trial and alternative drugs similarly affect the measured outcome.

The described framework is easily customizable for various diseases. Specifically, to study acute conditions, we can set the follow-
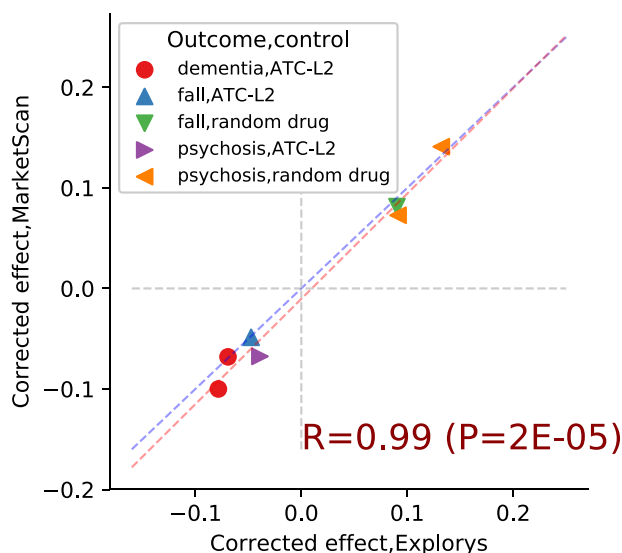
Figure 4. Estimated causal effects: Explorys vs MarketScan. Each point corresponds to a drug estimated to have a significant effect in both databases. Marker type represents the combination of trial outcome and control cohort; points in the first and third quadrant indicate harmful and beneficial effects, respectively. Red line is the least squares regression line, blue line is $y = x$.

of treatments and their duration and is expected to result in smaller treatment cohorts.

We note several additional directions for extending our framework. Currently, we estimate the average effect for the entire population, although effects may be heterogeneous with large differences across population strata. Identifying the sub-populations that benefit the most from each given drug (see Ozery-Flato et al.[48] for potential approaches) could focus drug development efforts. Other future directions include supporting time-varying confounders and treatments to better capture temporal causal trends, incorporating drug dosage in the analysis, and inspecting the effect of inactive drug ingredients.

CONCLUSION

We presented a flexible computational framework for high-throughput identification of drug repurposing candidates that efficiently emulates hundreds of RCTs from observational medical data to estimate the effect of on-market drugs on various disease outcomes. Naturally, the generated hypotheses require clinical analysis and experimental validation, but the significant agreement across databases and methodological approaches is encouraging. Notably, our framework may augment other *in silico* approaches[4] that leverage drug- or disease-related characteristics to identify promising drug repurposing candidates.

FUNDING

AUTHOR CONTRIBUTIONS

MOF designed the study, developed the framework, conducted the analyses, and wrote the manuscript; YG conceptualized the study; OS contributed to the framework and analysis; SR contributed to the framework; CY conceptualized and designed the study, contributed to the analysis, and wrote the manuscript. All authors reviewed, provided input, and accepted the submitted version.

SUPPLEMENTARY MATERIAL

Supplementary material is available at *Journal of the American Medical Informatics Association* online.

CONFLICT OF INTEREST STATEMENT

The authors have no competing interests to declare.

up period to few months or days. Our framework is also extendible, as its components can be configured to use alternative implementations to the ones described above. A central modifiable component is the causal effect estimation method. In the PD case study, we tested two distinct approaches for causal effect estimation: balancing weights and outcome prediction. A straightforward extension is to use doubly robust methods,[40] which combine the two previous approaches. We also used balancing weights for assessing whether treatment and control biases can be successfully eliminated (Steps 2–9 in Procedure 1). Approximately 6–18% of the tested drugs failed this assessment (see Table 1). Regression-based methods, such as Cox model, which were commonly used in previous drug-repurposing studies, do not allow one to determine whether identified biases have been eliminated.[26] We obtained balancing weights using the classical IPW method, which may suffer from large estimation variance and is sensitive to model misspecification. There are many alternative methods for IPW,[41–47] and each of these methods can be plugged into our framework. Using different inference methods allows the reader to evaluate the sensitivity of identified effects to modeling decisions, as suggested by Brookhart et al.[35] Alternative implementations to other algorithmic steps, for example, confounder selection, may provide an even more comprehensive evaluation of the obtained results and their robustness.

In the current study, we tested only individual active ingredients, corresponding to specific molecules. Laifenfeld et al[19] further inspected and analyzed two of the ingredients identified by our framework as beneficial for PD patients, namely rasagiline and zolpidem, and proposed plausible mechanistic explanations for the observed effects. Focusing on individual drugs may overlook significant effects shared by multiple similar molecules whose independent analysis lacks statistical power. To overcome this issue, we can define the set of tested drugs to be families of related molecules (eg, using the ATC drug classification system). Similarly, we may consider drug combinations to obtain insights on synergetic effects of molecules, though such analysis requires extending the definitions

REFERENCES

1. Langedijk J, Mantel-Teeuwisse AK, Slijkerman DS, *et al.* Drug repositioning and repurposing: terminology and definitions in literature. *Drug Discov Today* 2015; 20 (8): 1027–34.
2. Ashburn TT, Thor KB. Drug repositioning: identifying and developing new uses for existing drugs. *Nat Rev Drug Discov* 2004; 3 (8): 673–83.
3. Pushpakom S, Iorio F, Eyers PA, *et al.* Drug repurposing: progress, challenges and recommendations. *Nat Rev Drug Discov* October 2018; 18 (1): 41–58.
4. Hurle MR, Yang L, Xie Q, *et al.* Computational drug repositioning: from data to therapeutics. *Clin Pharmacol Ther* 2013; 93 (4): 335–41.
5. Li J, Zheng S, Chen B, *et al.* A survey of current trends in computational drug repositioning. *Brief Bioinform* 2016; 17 (1): 2–12.

6. Hodos RA, Kidd BA, Shameer K, et al. In silico methods for drug repurposing and pharmacology. *Wires Syst Biol Med* 2016; 8 (3): 186–210.

7. Chen B, Butte A. Leveraging big data to transform target selection and drug discovery. *Clin Pharmacol Ther* 2016; 99 (3): 285–97.

8. Paranjpe MD, Taubes A, Sirota M. Insights into computational drug repurposing for neurodegenerative disease. *Trends Pharmacol Sci* 2019; 40 (8): 565–76.

9. Sherman RE, Anderson SA, Dal Pan GJ, et al. Real-world evidence—what is it and what can it tell us? *N Engl J Med* 2016; 375 (23): 2293–7.

10. Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. *Nat Rev Genet* 2012; 13 (6): 395–405.

11. Xu H, Aldrich MC, Chen Q, et al. Validating drug repurposing signals using electronic health records: a case study of metformin associated with reduced cancer mortality. *J Am Med Inform Assoc* 2015; 22 (1): 179–91.

12. Paik H, Chung A-Y, Park H-C, et al. Repurpose terbutaline sulfate for amyotrophic lateral sclerosis using electronic medical records. *Sci Rep* 2015; 5 (1): 8580.

13. Kuang Z, Bao Y, Thomson J, et al. A Machine-learning based drug repurposing approach using baseline regularization. *Methods Mol Biol* 2019; 1903: 255–67.

14. Suchard MA, Zorych I, Simpson SE, et al. Empirical performance of the self-controlled case series design: lessons for developing a risk identification and analysis system. *Drug Saf* 2013; 36 (S1): 83–93.

15. Wu Y, Warner JL, Wang L, et al. Discovery of noncancer drug effects on survival in electronic health records of patients with cancer: a new paradigm for drug repurposing. *JCO Clin Cancer Inform* 2019; (3): 1–9.

16. Suchard MA, Schuemie MJ, Krumholz HM, et al. Comprehensive comparative effectiveness and safety of first-line antihypertensive drug classes: a systematic, multinational, large-scale analysis. *Lancet* 2019; 394 (10211): 1816–26.

17. Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. *Am J Epidemiol* 2016; 183 (8): 758–64.

18. Lang AE, Espay AJ. Disease modification in Parkinson's disease: current approaches, challenges, and future considerations: disease modification in PD. *Mov Disord* 2018; 33 (5): 660–77.

19. Laifenfeld D, Yanover C, Ozery-Flato M, et al. Emulated clinical trials from longitudinal real-world data efficiently identify candidates for neurological disease modification: examples from parkinson's disease. *medRxiv.* 2020. doi: 10.1101/2020.08.11.20171447.

20. Gomez Arevalo G, Jorge R, Garcia S, et al. Clinical and pharmacological differences in early- versus late-onset Parkinson's disease. *Mov Disord* 1997; 12 (3): 277–84.

21. Pagano G, Ferrara N, Brooks DJ, et al. Age at onset and Parkinson disease phenotype. *Neurology* 2016; 86 (15): 1400–7.

22. Hernan MA, RobinsJM. *Causal Inference*. Boca Raton: Chapman & Hall/CRC; 2019.

23. Ozery-Flato M, Yanover C, Gottlieb A, et al. Fast and efficient feature engineering for multi-cohort analysis of EHR data. *Stud Health Technol Inform* 2017; 235: 181–5.

24. Brookhart MA, Schneeweiss S, Rothman KJ, et al. Variable selection for propensity score models. *Am J Epidemiol* 2006; 163 (12): 1149–56.

25. Wasserman L. *All of Statistics: A Concise Course in Statistical Inference*. Berlin, Germany: Springer Science & Business Media; 2013.

26. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behav Res* 2011; 46 (3): 399–424.

27. Cole SR, Hernán MA. Constructing inverse probability weights for marginal structural models. *Am J Epidemiol* 2008; 168 (6): 656–64.

28. Gottlieb A, Yanover C, Cahan A, et al. Estimating the effects of second-line therapy for type 2 diabetes mellitus: retrospective cohort study. *BMJ Open Diab Res Care* 2017; 5 (1): e000435.

29. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: L. Erlbaum Associates; 1988.

30. Beaudoin FL, Gutman R, Merchant RC, et al. Persistent pain after motor vehicle collision: comparative effectiveness of opioids versus non-steroidal anti-inflammatory drugs prescribed from the emergency department—a propensity matched analysis. *Pain* 2017; 158 (2): 289–95.

31. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B Methodol*1995; 57 (1): 289–300.

32. Pearl J. Causal inference in statistics: an overview. *Statist Surv* 2009; 3: 96–146.

33. Hernán MA, Hsu J, Healy B. Data science is science's second chance to get causal inference right: a classification of data science tasks. *Chance* 2019; 32 (1): 42–9.

34. Hersh WR, Weiner MG, Embi PJ, et al. Caveats for the use of operational electronic health record data in comparative effectiveness research. *Med Care* 2013; 51 (8 Suppl 3): S30–7.

35. Brookhart M, Stürmer T, Glynn R, et al. Confounding control in healthcare database research: challenges and potential approaches. *Med Care* 2010; 48: S114–20.

36. Pearl J. On a Class of Bias-Amplifying Variables that Endanger Effect Estimates. *arXiv : 12033503 [cs, stat]* 2012. http://arxiv.org/abs/1203.3503.

37. Vansteelandt S, Bekaert M, Claeskens G. On model selection and model misspecification in causal inference. *Stat Methods Med Res* 2012; 21 (1): 7–30.

38. Ertefaie A, Asgharian M, Stephens DA. Variable selection in causal inference using a simultaneous penalization method. *J Causal Inference* 2018; 6 (1) doi: 10.1515/jci-2017-0010.

39. Ryan PB, Schuemie MJ, Gruber S, et al. Empirical performance of a new user cohort method: lessons for developing a risk identification and analysis system. *Drug Saf* 2013; 36 (S1): 59–72.

40. Robins J, Sued M, Lei-Gomez Q, et al. Comment: performance of double-robust estimators when "inverse probability" weights are highly variable. *Stat Sci* 2007; 22 (4): 544–59.

41. Hainmueller J. Entropy balancing for causal effects: a multivariate reweighting method to produce balanced samples in observational studies. *Polit Anal* 2012; 20 (1): 25–46.

42. Imai K, Ratkovic M. Covariate balancing propensity score. *J R Stat Soc B* 2014; 76 (1): 243–63.

43. Zubizarreta JR. Stable weights that balance covariates for estimation with incomplete outcome data. *J Am Stat Assoc* 2015; 110 (511): 910–22.

44. Chan KCG, Yam SCP, Zhang Z. Globally efficient non-parametric inference of average treatment effects by empirical balancing calibration weighting. *J R Stat Soc B* 2016; 78 (3): 673–700.

45. Kallus N. Balanced policy evaluation and learning. In: Bengio S, Wallach H, Larochelle H, et al., eds. *Advances in Neural Information Processing Systems 31*. Curran Associates, Inc.; 2018: 8895–8906. http://papers.nips.cc/paper/8105-balanced-policy-evaluation-and-learning.pdf.

46. Zhao Q. Covariate balancing propensity score by tailored loss functions. *Ann Statist* 2019; 47 (2): 965–93.

47. Ozery-Flato M, Thodoroff P, El-Hay T. Adversarial Balancing for Causal Inference. *arXiv : 181007406 [cs, stat]* 2018. http://arxiv.org/abs/1810.07406.

48. Ozery-Flato M, El-Hay T, Aharonov R, et al. Characterizing subpopulations with better response to treatment using observational data—an Epilepsy case study. *bioRxiv* 2018. doi:10.1101/290585.